2000년 한국정보처리학회 추계 학술발표논문집 제7권 제2호

다중 경로 버퍼를 이용한 오류허용 ATM 스위치

신원철°, 손동욱, 손유익
계명대학교 컴퓨터공학전공
synn@jinri.kmu.ac.kr, psalm8@hcc.ac.kr, yeson@kmu.ac.kr

A Fault-Tolerant ATM Switch using Multiple-Path Buffers

Won-Chul Synn, Dong-Wuk Son, Yoo-Ek Son
Department of Computer Science Keimyung University

요약

ATM 스위치로 많이 이용되는 다단계 상호연결 네트워크(MIN)는 self-routing 및 one-to-one 연결 특성을 가진 블록킹 네트워크로써 셀 전송시 충돌이 일어날 수 있다. 따라서 버퍼를 갖는 스위치를 고려하게 된다. 본 논문에서는 스위치의 내부에 다중경로를 제공하는 입력버퍼를 이용하여 각 스위치의 입력포트에서 출력포트로의 경로를 확장시킨 스위치 구조 및 네트워크의 성능 향상에 대하여 언급한다. 이를 위해 네트워크의 stage간 상호연결 패턴이 buddy 및 constrained reachability 특성에 따른 경로설정 구조를 기본으로 이용한다. 그리고 입력버퍼 스위치 구조의 문제점인 HOL 블록킹의 방지 및 오류허용 기능을 향상시킬 수 있는 다중경로 버퍼를 갖는 ATM 스위치 구조를 제안하고, 시뮬레이션을 통해 그 성능을 분석한다.

1. 서론

다단계 상호연결 네트워크는 메모리 모듈과 프로세서의 상호 연결을 위한 병렬 컴퓨터구조 및 초고속통신시스템에서의 효과적인 교환구조로 인식되면서 많은 연구들이 진행되어 왔다.

입력버퍼 구조의 스위치소자들로 구성되는 네트워크의 경우, 그 특성상 단일경로를 갖는 문제점으로 인하여 블록킹의 가능성이 높고 이로 인해 셀의 손실을 가져올 수 있다. 블록킹에 따른 성능저하를 방지하기 위한 방법으로는 버퍼링, 정렬 네트워크, 다중경로를 이용하는 등 다양한 기법이 있지만, 하드웨어의 복잡성, 제어를 위한 부가적인 요구사항, 성능 대 비용 등의 이유로 일반적으로 버퍼링 기법이 사용된다.[3]

입력 버퍼링의 경우, 만약 FIFO 입력버퍼가 각 입력포트에서 셀을 전송하기 위해 사용된다면 각 버퍼에서 입력되는 첫 번째 셀 만이 앞으로 진행됨으로써 결과적으로 HOL(Head Of Line) 블록킹이 발생하게 된다. 균일트래픽(uniform traffic)에서 입력버퍼 스위치의 최대 처리량은 이로 인해 약 58.6% 정도 된다는 점은 이미 알려진 바와 같다[5]. 더욱이 버스트 트래픽(burst traffic)인 경우 최대 처리량은 50%로 더욱 낮아진다. 이러한 경우 각 입력 부하에 대한 셀 손실률은 버퍼의 크기와 관계없이 높아지게 되므로 이러한 셀 손실을 막기 위해서는 입력버퍼 스위치의 구조와 각 stage간의 링크를 확장시킴으로써 가능하게 할 수 있다.[7]

일반적으로 ATM 스위치는 버퍼링 방식 및 내부 스위칭, 스위칭 구조에 따라 구분되는데[3], 본 논문은 다단계 상호연결 네트워크의 일종인 베이스라인 네트워크를 대상으로, 스위치소자 내부에 다중경로를 제공하는 입력버퍼를 사용한 구조를 제안함으로써 입력버퍼의 문제점인 HOL 블록킹의 방지 및 오류허용 기능을 향상시키고자 하였다. 또한 시뮬레이션을 통하여 외부 입력버퍼 구조의 베이스라인 네트워크 및 내부 입력버퍼 네트워크[2]와의 셀 손실률 및 지연, 그리고 처리량 등의 성능을 비교하고 이를 통해 제안된 구조의 성능향상에 대해서 설명하고자 한다.

2. MIN 특성

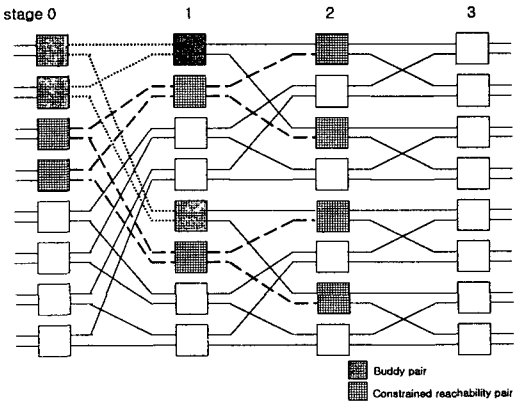
본 논문에서는 다단계 상호 연결 네트워크의 대표적인 구조중 하나인 베이스라인 네트워크를 적용대상으로 다룬다. 베이스라인 네트워크의 토폴로지는 재귀적 방법으로 구성된다. 즉, 첫 번째 stage는 N×N 블럭을 포함하고, 두 번째 stage는 (N/2)×(N/2) 하위블럭을 포함하는 등 (log_2N-1)회 반복된다.

베이스라인 네트워크에서는 물리적 이름과 논리적 이름은 같으며 stage 레이블 표현은 0에서 (log_2N-1)까지 연속적이다. 링크레벨의 표현 또한 0에서 (log_2N-1)까지 연속적으로 표현된다. 스위치소자 l은 (log_2N-1)이 되고 이진표현으로는 $p_{l-1}\cdots p_1$이 된다. 각 레벨 내의 링크는 $p_l p_{l-1}\cdots p_0$, $p_l p_{l-1}\cdots p_1$까지는 동일하지만 p_0=0이면 링크가 스위치소자의 상위 출력포트에 연결되고, p_0=1이면 하위 출력포트에 연결된다. *stage[i]*에서의 스위치소자의 물리적 이름은 $(p_l p_{l-1}\cdots p_1)_i$이고, level i에서 링크의 물리적 이름은 $(p_l p_{l-1}\cdots p_0)_i$이 된다. 이에 따른 베이스라인 네트워크의 상호 연결은 다음과 같이 표현된다.

$$\beta_i^0[(p_l p_{l-1}\cdots p_1)_i] = (p_l\cdots p_{l-i+1}0p_{l-i}\cdots p_2)_{i+1}$$
for link $(p_l p_{l-1}\cdots p_1 0)_{i+1}$, $0 \le i < l$

$$\beta_i^1[(p_l p_{l-1}\cdots p_1)_i] = (p_l\cdots p_{l-i+1}1p_{l-i}\cdots p_2)_{i+1}$$
for link $(p_l p_{l-1}\cdots p_1 1)_{i+1}$, $0 \le i < l$

이러한 상호 연결 특성으로 인해 베이스라인 네트워크는 buddy 및 constrained reachability의 두 가지 특성에 의한 상호연결 패턴을 가진다.[4] Buddy 특성은, 만약 *stage[i]*에서 $SE[j_i]$가 $SE[l_{i+1}]$와 $SE[m_{i+1}]$과 연결된다면, 두 스위치소자는 *stage[i]*에서 동일한 $SE[k_i]$와 연결된다는 것이다. 다시 말해서 인접한 stage의 스위치소자들은 항상 각각 쌍을 이루어 상호 연결된다. Buddy 속성을 적용시켜 보면 *stage[i]*에서 모든 스위치소자들은 *stage[i+k]*에서 같은 크기의 스위치소자 쌍들과 연결된다는 것을 알 수 있다.



[그림 1] Buddy 및 reachability 특성에 따른 연결구조

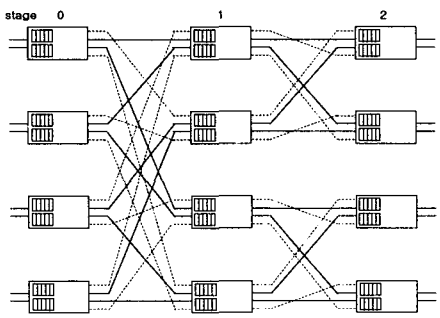
Reachability 특성은 *stage[i]*에서의 스위치소자에 의해 *stage[i+k]*에 연결된 2^k개의 스위치소자는 *stage[i]*에서 정확히 2^{k-1}개의 서로 다른 스위치소자들과 연결된다. 이것은 연결된 스위치소자 상호간의 집합을 찾아내기 위한 buddy 특성으로부터 나온 것이다. 즉, 하나의 스위치소자는 *stage[i+1]*에서 2개의 스위치소자와 연결되고, *stage[i+k-1]*에서는 2^{k-1}개의 스위치소자와 경로가 설정된다. 그러므로 *stage[i+k]*에서 2^k개의 경로선택이 가능하다. [그림 1]은 위에서 언급한 두 가지 특성을 나타낸 것이다. 그림에서 보는 바와 같이 *stage[0]*과 *stage[1]*의 두 쌍의 buddy 스위치소자는 *stage[0]*에서 *stage[2]*을 통해 네 개의 스위치소자에서 reachability 특성을 가진다.

본 논문은 이 두 가지 속성을 이용하여 *stage[i]*에서 *stage[i+1]*로 연결되는 링크를 확장시켜 라우팅 가능 경로 수를 증가시킴으로써 HOL 블록킹으로 인한 셀 손실을 줄이고 버퍼의 효율을 향상시키고자 하였다.

3. 다중 경로 구조

베이스라인 네트워크에서의 셀 라우팅은 입출력간 단일 경로로 인해 하나의 경로만이 존재한다. 즉, *stage[X]*에서 라우팅 태그는 n-X+1번째 목적지 주소의 비트가 된다.

제안된 네트워크에서의 각 스위치소자는 두 개의 여분 입력 및 출력링크와 두 개의 버퍼모듈로 구성되어 있다.(*stage[0], [n-1]*은 제외) 스위치소자에 추가된 두 개의 링크는 두 개의 stage사이의 경로를 확장시켜 다음 stage에서의 버퍼가 블록킹 혹은 링크나 스위치소자에서 오류가 생겼을 경우 redundant 링크로 사용된다. [그림 2]는 $2^3\times2^3$ 크기의 다중 경로 구조를 보여주고 있다. 그림에서 알 수 있듯이 크기가 N인 베이스라인 네트워크는 순환적인 구조를 가지고 있으며, $\sum 2^i$(i는 stage)개의 하위 네트워크로 분할이 된다.



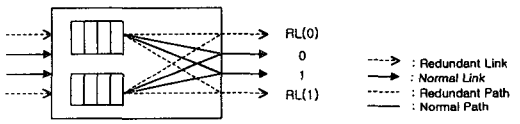
[그림 2] $2^3\times2^3$ 크기의 다중 경로 구조

분할된 각 하위 네트워크로 라우팅이 될 경우 경로 선택은 병렬적으로 선택되어지며, 이러한 특성으로

인해 redundant 경로가 결정되게 된다.

여기서 제시한 네트워크에서 사용되는 스위치소자의 출력링크는 각각 0, 1, RL(0), RL(1) 이다. 여기서 RL(0)과 RL(1)은 각각 라우팅 비트 0과 1의 redundant 링크이며, *stage[X]*$(0 \le X \le n-2)$에서 스위치소자의 각 버퍼모듈은 각 출력포트에 연결된 normal 경로와 redundant 경로로 사용된다. 상위 및 하위 버퍼모듈의 normal 경로는 normal 링크와 연결되며, redundant 경로는 redundant 링크와 연결된다. 그리고 각 스위치소자에 입력되는 두 개의 링크는 서로 다른 곳으로부터 연결되게 된다. 즉, redundant 링크와 normal 링크가 각각 분리되지 않고 하나의 버퍼에서 같이 처리되기 때문에 버퍼의 이용률은 각각 따로 처리할 때보다 높아진다.

[그림 3]은 제안된 네트워크 구조에서 사용되는 네 개의 내부링크를 가진 스위치소자를 보여주고 있다.



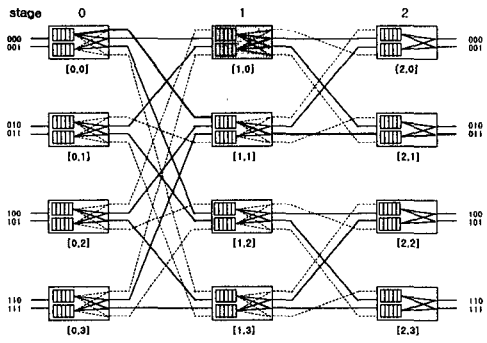
[그림 3] 네 개의 내부링크를 가진 스위치소자

이 구조에서는 두 개의 경로(normal 및 redundant 경로)가 각 스위치소자에 존재하게 된다. 그러므로 *stage[X]*$(0 \le X \le n-2)$에서는 연결될 스위치소자 또는 링크에 블록킹이나 오류가 없다면 각 입력노드는 normal 경로를 통해 셀을 전송한다. 그러나 오류가 있을 경우 입력노드는 redundant 경로를 사용하게 된다. *stage*$[n-1]$에서는 normal 경로만이 존재하며 베이스라인 네트워크 라우팅 규칙에 따른다.

셀의 수락은 첫 stage부터 마지막 stage에 걸쳐 적용된다. 다음 stage에서 수락하게 되면 스위치소자는 버퍼모듈의 버퍼가 얼마만큼이나 사용가능한지, 그리고 다음 stage로의 링크 및 스위치소자의 오류유무를 알게 되고, 입력되는 셀의 수락을 결정하게 되면 셀 수락정보를 뒤로 전파한다.

수락된 셀은 항상 전송될 버퍼모듈의 마지막 부분으로 오게 된다. 수락이 거부된 셀은 버퍼모듈의 앞부분에서 대기하고 다음 stage 사이클에 전송된다. 그리고 스위치소자 혹은 링크의 오류여부와 관계없이 입출력 라우팅 경로가 유일하게 존재하기 때문에 셀의 순서는 유지된다. 단 오류가 발생한 스위치소자 혹은 링크가 발견되어 새로운 경로가 설정되는 일시적인 기간동안은 제외된다.

[그림 4]는 입력 0에서 출력 3으로 셀 전송이 일어날 경우, *SE[1,0]*가 오류 또는 버퍼에서 블록킹이 발생할 때 redundant 경로를 통해 셀이 라우팅되는 경로를 보여주고 있다.

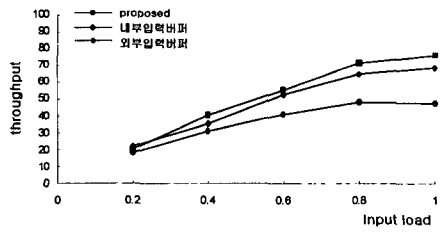


[그림 4] 오류 발생시 라우팅

4. 성능평가

본 논문에서는 네트워크의 성능평가를 위해 이산적인 모델링 방식으로 시뮬레이션 하였으며, 셀 손실률(cell loss), 처리량(throughput), 셀 지연(cell delay)면에서 성능을 평가하였다.

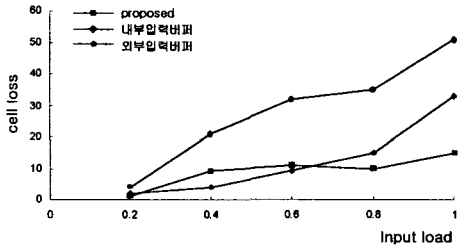
시뮬레이션 모델링을 위한 조건으로는, 각 스위치소자는 2×2의 크기(본 논문에서 제시한 스위치소자 제외)이며, 버퍼의 크기는 4, 네트워크 크기는 8×8로 하였다. 그리고 고정된 크기의 셀을 전송하는 ATM 스위치의 환경을 반영하기 위해 스위치의 동작은 동기적이며, 각 입력에 도착하는 셀의 시간간격은 exponential 분포를 따른다.[1] 그리고 마지막 stage에서의 블록킹은 고려하지 않으며 출력링크의 속도는 스위치 내부의 링크속도보다 빠르다고 가정한다. 각 입력포트의 입력부하는 동일하며 입력포트에 도착하는 셀의 시간간격과 스위칭 시간간격이 같을 경우 입력부하는 1이 된다. 마지막으로 셀의 라우팅 경로는 임의로 결정된다. 각 네트워크의 처리량은 매 주기마다 출력되는 셀의 개수로 시뮬레이션에서는 일정한 시간 안에 실제로 처리한 셀 수의 합으로 정의하였다.



[그림 5] 입력부하에 따른 처리량

[그림 5]는 입력부하에 따른 각 네트워크의 처리량

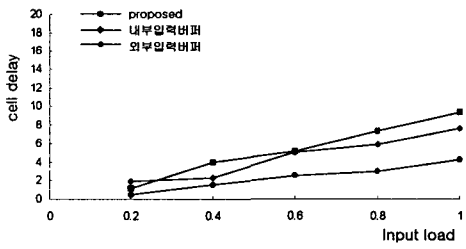
을 나타내고 있다. 그림에서 보면 세 가지 네트워크 모델이 모두 선형적인 증가를 보이고 있는데, 본 논문에서 제시한 네트워크 구조의 처리량이 다른 두 가지 네트워크에 비해 높은 처리율을 얻을 수 있다는 것을 알 수 있다. 이것은 버퍼에서 다중경로를 제공하여 다음 stage에서 블록킹이 발생하는 경우 redundant 경로로 셀이 진행되기 때문이다.



[그림 6] 입력부하에 따른 셀 손실의 변화

[그림 6]은 셀 손실에 대한 성능비교로써, 셀 손실은 스위치소자에 입력된 총 셀의 개수에 대해 출력포트로 출력되지 못하고 손실되는 셀의 비율로 정의하였다. 여기서도 본 논문에서 제시한 네트워크 구조의 손실은 다른 두 가지 네트워크 구조가 입력부하가 높아짐에 따라 급격히 높아지는데 반해, 천천히 상승하고 있는 모습을 보이고 있다. 이것은 입력부하가 늘어나면 셀의 스위칭 시간과 셀이 생성되는 시간이 같게 되어 전송되는 셀들이 블록킹으로 인해 순차적으로 진행하지 못하기 때문이다. 그래서 유일한 경로만을 가진 다른 네트워크의 경우 블록킹을 피할 수 없어 셀 손실은 더욱 커지게 된다.

마지막으로 셀 지연에 대한 변화이다. 셀 지연은 일반적으로 스위치 지연(switch delay)과 버퍼지연(buffered delay)으로 나뉘게 되는데[6], 여기서는 각 셀들의 버퍼지연으로 평가하였다



[그림 7] 입력부하에 따른 셀 지연의 변화

셀 지연은 입력포트로 들어온 셀이 출력포트로 출력될 때까지 각 셀들이 버퍼에 대기한 평균대기시간(average waiting time)이 된다. [그림 7]에서 세 가지 네트워크 구조의 셀 지연은 확연한 차이를 보이지 않고 거의 비슷한 변화를 보이고 있음을 알 수 있다.

5. 결론

본 논문은 다단계 상호연결 네트워크(MIN)의 대표적 구조인 베이스라인 네트워크를 대상으로, 입력버퍼 스위치 구조의 문제점인 HOL 블록킹의 방지 및 오류허용 기능을 향상시킬 수 있는 다중경로 버퍼를 갖는 ATM 스위치 구조를 제안하고, 시뮬레이션을 통해 그 성능을 분석한다. 이를 위해 buddy 및 reachability 특성의 상호연결 패턴의 구조적 특성을 이용하고, 또한 스위치소자에 내부경로를 확장시켰다. 분석을 위해 스위치 내의 HOL 블록킹으로 인한 셀 손실, 링크에 대한 오류허용 기능 및 성능 향상에 대해서 언급하였으며 시뮬레이션을 통해 셀 손실 및 지연, 그리고 처리율에 대하여 평가하였다.

시뮬레이션 결과, 본 연구에서 제안된 모델은 기존의 외부 입력버퍼 구조와 내부 입력버퍼 구조에 비해 높은 처리율과 낮은 셀 지연을 보임으로써 높은 성능을 나타내고 있으며, 특히 입력부하가 높을 경우 기존의 모델보다 더 나은 성능을 가지고 있음을 알 수 있다.

그러나 입력부하에 따른 전체적인 성능의 평가뿐만 아니라 차후, 버퍼의 크기, 네트워크 크기에 따른 성능평가와 오류허용에 대한 검증 또한 필요할 것으로 생각된다.

참고문헌

[1] Thomas H. Theimer, Erwin P. Rathgeb, Manfred N. Huber, "Performance Analysis of Buffered Banyan Networks," IEEE Comm, p.413-421, Feb 1991.

[2] Jianxu Ding, Laxmi N. Bhuyan "Finite Buffer Analysis of Multistage Interconnection Networks," IEEE Trans, p.23-247, Vol.43, No.2, Feb. 1994..

[3] H. Kim, A. Ahmad, C. Oh, K. Kim, "Performance Comparison High-Speed Input-Buffered ATM Switches," Proc. IEEE ATM '97 Workshop, Lisbon, Portugal, p.505-513, May 1997.

[4] Achille Pattavina "Switching Theory : Architecture and Performance in Broadband ATM Networks," Wiley, 1997.

[5] I. I. Makhamreh "Throughput Analysis of Input-Buffered ATM Switch," IEEE Proc. Comm, Vol.145, No.1, Feb 1998.

[6] Nick McKeown, Tomas E. Anderson "A quantitative comparison of iterative scheduling algorithm for input-queued switches" Elsevier Science B.V. Computer Networks and ISDN Systems30, p.2309-2326, 1998.

[7] B. Kraimeche "Design and analysis of the Stacked Banyan ATM switch fabric," Elsevier Science B.V. Computer Networks32, p.171-184, 2000.